DOCUMENT RESUME

ED 073 787 LI 004 234

AUTHOR	Wright, Kieth C.
TITLE	Trends in Modern Subject Analysis with Reference to Text Derivative Indexing and Abstracting Methods: The State of the Art.
INSTITUTION	ERIC Clearinghouse on Library and Information Sciences, Washington, D.C.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	72
NOTE	19p.; (135 References)
AVAILABLE FROM	Information-Part 2, Science Associates/ International, 23 East 26th Street, New York, New York 10010 (no price quoted)
JOURNAL CIT	Information-Part 2, September-October 1972
EDRS PRICE	MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS	*Abstracting; *Automatic Indexing; *Automation; Classification; Electronic Data Processing; *Indexing; *Information Processing; Literature Reviews; State of the Art Reviews

ABSTRACT

This paper briefly reviews the information explosion of the last thirty years and the various attempts made to organize that information in new ways. Section B offers a brief historic review of modern classification and subject heading theory. Section C reviews the literature of automatic indexing, automatic abstracting, and automatic classification. The problems of large file organization, word meanings, and the limitation of such "automatic" methods are discussed. Section D sums up the state of the art in automatic indexing by concluding that human intellectual effort is still required in indexing. The computer is viewed as a valuable assistant in that intellectual effort and the wide variety of computer applications to indexing work is summarized. (Author)

ED 073787

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY

Science Associates/Int'l

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

FILMED FROM BEST AVAILABLE COPY

TRENDS IN MODERN SUBJECT ANALYSIS WITH REFERENCE TO TEXT DERIVATIVE INDEXING AND ABSTRACTING METHODS: THE STATE OF THE ART

by

Kieth C. Wright

The Edward Miner Gallaudet Memorial Library
Gallaudet College

Published in cooperation with the

ERIC Clearinghouse on Library and Information Sciences
1140 Connecticut Avenue, N.W., Suite 804
Washington, D. C. 20036

LI 004 234

This publication was prepared pursuant to a contract with the National Institute of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not, therefore, necessarily represent official National Institute of Education position or policy.

TRENDS IN MODERN SUBJECT ANALYSIS WITH PARTICULAR REFERENCE TO TEXT DERIVATIVE INDEXING AND ABSTRACTING METHODS: THE STATE OF THE ART

A. *INTRODUCTION*

This paper briefly reviews the information explosion of the last thirty years and the various attempts made to organize that information in new ways. Section B offers a brief historic review of modern classification and subject heading theory.

Section C reviews the literature of automatic indexing, automatic abstracting, and automatic classification. The problems of large file organization, word meanings, and the limitation of such "automatic" methods are discussed.

Section D sums up the state of the art in automatic indexing by concluding that human intellectual effort is still required in indexing. The computer is viewed as a valuable assistant in that intellectual effort and the wide variety of computer applications to indexing work is summarized.

B. *THE "INFORMATION EXPLOSION" AND ATTEMPTS TO MEET IT*

"The cost of manual classification and abstracting of all the articles in the world's hundred-thousand technical periodicals would be fantastic. The practicality of carrying it out in a coordinated timely way by manual methods is unrealizable. There is also a pressing need to extend the coverage of a myriad of unpublished working papers. Hence, there is an utter necessity for automatic indexing, abstracting and summaries made by electronic data processing."[1]

The "information explosion," alluded to above, and the growing difficulty experienced in handling that volume of information in an efficient way to serve particular users have combined to create an increased interest in unconventional, often machine-aided, information retrieval systems.[2] Lilley has described the radical changes in information flow during the past fifty years and related these changes to the rise of nonhierarchical classification ideas and new subject analysis techniques.[3]

Bourne illustrates the magnitude of the paper problem with these examples: (1) The federal government produces twenty-five billion pieces of paper a year. There is now enough research to fill 7.5 Pentagons with file cabinets at a cost of over four billion dollars a year. (2) Military engineering drawings and documents cost two billion dollars a year and yield six million drawings which must be added to the fifty million already on file. (3) There are 30,000 technical journal publications with over two million articles annually.[4] Obviously all of these pieces of paper do not contain new information as anyone who currently reads the journals, abstracts and microfilms of a field will quickly discover.[5]

---

[1]B. F. Cheydleur. "Information Retrieval - - - 1966." DATAMATION, 7:10, (October, 1961), pp. 21-25. Quote is from p. 21.

[2]Jesse H. Shera. "Trends in Subject Analysis Practice, 1950 to Today." American Library Association, Preconference Institute, The Subject Analysis of Library Materials, Atlantic City, NJ, June 19-21, 1969. (Typescript).

[3]Oliver L. Lilley. "Terminology, Form, Specificity and the Syndetic Structure of Subject Headings for English Literature." (D. L. S. Thesis, Columbia University School of Library Service), 1959, pp. 14ff.; Jack C. Morris. "The Duality Concept in Subject Analysis." (Oak Ridge: Oak Ridge National Laboratories, 1953), 46 p. gives a similar assessment of these changes.

[4]Charles P. Bourne, ed. METHODS OF INFORMATION HANDLING. Wiley, 1963, 241 p. Bourne's figure on periodicals is in dispute; K. P. Barr in his article, "The Estimate of the Number of Currently Available Scientific and Technical Periodical." JOURNAL OF DOCUMENTATION, 23:2 (June, 1967), pp. 110-116, sets the total at about 26,000 based on the experience of the British National Lending Library for Science and Technology. On the other hand, C. M. Gottschalk and W. F. Desmond in their "Worldwide Census of Scientific and Technical Serials." AMERICAN DOCUMENTATION, 14:3 (July, 1963), p. 188, set the figure at 35,000.

[5]A point stressed by Jesse H. Shera in his "The Sociological Relationship of Information Science." JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE, (Henceforth JASIS), 22:2, (March-April, 1971), pp. 76-80; "...but it is neither an explosion nor information...it is a paper explosion; it is not knowledge, but paper that is increasing at an exponential rate." p. 78. See also, Y. Bar Hillel, "Is Information Retrieval Approaching a Crisis?" AMERICAN DOCUMENTATION, 14, (1963), pp. 95-98.

As new information techniques arose to meet new information demands, new types of alphabetically arranged indexing aids such as thesauri were created.[6] Computer software programs were developed to supplement human efforts in information retrieval. These developments once again focused attention on the lack of a consistent, comprehensive code for subject heading creation and ordering.

Harris has considered the computer implications of rigorous definition in subject analysis. She found that the theory of subject headings had not greatly advanced since Cutter.[7] Pollard and Bradford pointed out over forty years ago that alphabetical subject indexes created problems because of their hidden, often unacknowledged, classification schemes.[8] Richmond investigated the concealed classification in the Library of Congress Subject Heading List under "cats." She found that much inconsistency seemed to be related to an unawareness of a classificatory scheme in the listing.[9] Daily's exhaustive study of the same subject heading list showed that once the words of a subject heading were chosen, the form of expression of that heading was largely determined by the need to fit the heading form into the existing structure of headings.[10] Frarey found that the emphasis during the 1940s shifted from "cataloger's choice" in headings to the use of common terms in the Library of Congress list. This shift caused an increasing complexity in the form of the subject heading list.[11]

Since alphabetical lists with subdivisions are often based on some type of classification, several authors have proposed that subarrangement or ordering of heading words or elements be made explicit by designating aspects formally. Kaiser developed a scheme for entry subdivision based on "concrete" and "process" aspects of a subject.[12] Prevost advocated a "noun rule" for the order of words within a heading.[13] Metcalf proposed that specificity of subject be absolute with relation to the object of study, and that specification of the aspects from which that object is studied be limited by the topic and the needs of the particular library.[14]

Modern exposition of "facet analysis" was begun by S.R. Ranganathan who applied his now famous principles of "Personality," "Matter," "Energy," "Space," and "Time" to indexable matter in order to create indexing access. He then developed a notational system utilizing the "colon" for linking these facets to one another.[15]

---

[6]Arthur L. Korotkin, Lawrence H. Oliver, and D. R. Burgis. INDEXING AIDS, PROCEDURES AND DEVICES, Report No. RADC-TR-64-582, (Bethesda, Maryland: General Electric Company, Information Systems Operation, April, 1965), AD 616 342, 110 p.; and Charles L. Bernier, "Indexing and Thesauri." SPECIAL LIBRARIES, 59 (February, 1968), pp. 98-103. Both publications survey developments and use of indexing aids.

[7]Jessica L. Harris. SUBJECT ANALYSIS; COMPUTER IMPLICATIONS OF RIGOROUS DEFINITIONS. Scarecrow, (1970), 279 p.; see especially page 13.

[8]A. F. C. Pollard and S. C. Bradford. "The Inadequacy of Alphabetical Subject Index." ASLIB PROCEEDINGS OF THE SEVENTH INTERNATIONAL CONFERENCE. (1930), pp. 39-45.

[9]Phyllis A. Richmond. "Cats: An Example of Concealed Classification in Subject Headings." LIBRARY RESOURCES AND TECHNICAL SERVICES, 3:2, (Spring, 1959), pp. 102-112.

[10]J. E. Daily. "The Grammer of Subject Headings." (D. L. S. Thesis, Columbia University School of Library Service, 1957) p. (unpublished), 222 p.

[11]Carlyle J. Frarey. "Subject Heading Revision by the Library of Congress." (Masters Paper, Columbia University School of Library Service, 1951), 97 p.; and his "A History of Subject Cataloging Principles and Practices in the United States, 1850-1954." (D. L. S. study in progress, Columbia University School of Library Service).

[12]J. Kaiser. SYSTEMATIC INDEXING, Pitman, (1911).

[13]Marie Louise Prevost. "Approach to Theory and Method in General Subject Headings." LIBRARY QUARTERLY, 16:2 (April, 1946), pp. 140-151.

[14]John W. Metcalf. INFORMATION INDEXING AND SUBJECT CATALOGUING. Scarecrow, (1957), 338 p.

[15]See his PROLEGOMENA TO LIBRARY CLASSIFICATION. Madras Library Association, (1937), Second Edition, Library Association, UK, (1957), 487 p.; and his "Subject Headings and Facet Analysis." JOURNAL OF DOCUMENTATION, 20:3 (September, 1964), pp. 109-119.

Later developments of the Colon classification and facet analysis indicate that application of facet principles is no easy matter, but must be carefully guided by a complex set of rules. The British Classification Research Group has been the major champion of facet classification research applications in England.[16]

Farradane has stressed the primacy of relationships between concepts and developed a system of relational "operators" used to connect concepts into sets called "analets."[17] Coates developed a formula for dealing with the multiple relationships between the various parts of a heading and for determining the order or words or elements in a heading.[18] Schemes such as those discussed above have not been extensively used in the United States. In 1960, Tauber and Lilley developed a faceted classification and index for a proposed Educational Media Research Information Service.[19] More recently Barhydt and Schmidt produced a thesaurus based on an analysis of facets and intended for use within the ERIC system.[20]

C. *AUTOMATIC ABSTRACTING, INDEXING, AND CLASSIFICATION*

While nonhierarchical classification schemes and theories were being developed, other researchers sought to automate the process of subject heading selection and ordering. Closely related research was undertaken in the areas of creating abstracts (or "extracts") automatically and classifying documents through automatic processes.

Prywes and Litofsky point out that the main justifications for all automatic processing of information are, "cost, personnel availability and service quality problems."[21] In 1967, Borko reviewed the types of automatic indexing and classification developed up to that time. He noted that the major difficulty was not in the counting and correlating of the characteristics of written language, but rather the problem was centered on what measures or counts are to be used in selecting actual terms for indexing. A basic assumption of all automatic indexing which utilizes counting or statistical methods is that, excluding function words like articles, conjunctions, and prepositions, "the more frequently a word is used in a document the more likely it is a significant indicator of subject matter."[22]

Doyle summarized developments in automatic classification; he observed that the increasing size of information stores on magnetic tape made some kind of organization of information essential in order to avoid costly sequential searches.[23] Stevens has produced an updating of her 1965 state of the art report on automatic

---

[16]Classification Research Group. "Need for a Faceted Classification as a Basis for all Methods of Information Retrieval." LIBRARY ASSOCIATION RECORD, 57:7 (July, 1955), pp. 262-268; see also "CRG Bulletin #9." JOURNAL OF DOCUMENTATION, 12:4 (December, 1956), pp. 227-230 and 273-298.

[17]J. E. L. Farradane. "A Scientific Theory of Classification and Indexing." JOURNAL OF DOCUMENTATION, 6:2 (June, 1950), pp. 83-99.

[18]E. J. Coates. SUBJECT CATALOGUES: HEADINGS AND STRUCTURE, Library Association, UK, (1960), 186 p.

[19]Maurice F. Tauber and Oliver L. Lilley. FEASIBILITY STUDY REGARDING THE ESTABLISHMENT OF AN EDUCATIONAL MEDIA RESEARCH INFORMATION SERVICE. Columbia University School of Library Service, (1960), 235 p.

[20]Gordon C. Barhydt and Charles T. Schmidt. INFORMATION RETRIEVAL THESAURUS FOR EDUCATION. Case Western Reserve University, (1970), 131 p.

[21]Noah S. Prywes and Barry Litofsky. "All-Automatic Processing for a Large Library." SPRING JOINT COMPUTER CONFERENCE, 36, May 5-7, 1970. AFIPS Press, (1970), pp. 323-333; quote is from page 323; John O'Conner stresses the same factors in his article, "Some Remarks on Mechanized Indexing, and Some Small Scale Results." MACHINE INDEXING: PROGRESS AND PROBLEMS, papers at the third International Institute on Information Storage and Retrieval, February 13-17, 1961, American University, (1962), 354 p. pp. 266-279.

[22]Harold Borko. "Indexing and Classification," in AUTOMATED LANGUAGE PROCESSING, edited by H. Borko. Wiley, (1967), pp. 99-125; quote is from p. 100.

[23]Lauren B. Doyle. "Is Automatic Classification a Reasonable Application of the Statistical Analysis of Text?" JOURNAL OF THE ASSOCIATION OF COMPUTING MACHINERY, (Henceforth JACM), 12:4 (October, 1965), pp. 473-489.

indexing.[24] The original report maintained that automatic indexing could increase the speed of indexing, increase the ease and speed of reindexing and reclassification, and increase the consistency of indexing efforts through consistent machine processes. The second report gives examples of small sample demonstrations which have shown the technical feasibility of statistical association methods for indexing. Such methods utilize the computer to count and calculate correlations among words and/or documents. Stevens stresses the need for extensive computer processing and analyses of large amounts of textual material in various subject fields similar to that done by Dennis in the legal field.[25]

More recently Batty has reviewed the last ten years of work in his "Automatic Generation of Index Languages."[26] He found that the basic statistical processes used in all of this work derived from a calculation based on the frequency of terms or words and their cooccurrences. Further he stresses that, apart from Stiles,[27] all of the work has been done with very small samples and that Doyle[28] has stressed the fantastic increase in costs when such techniques are applied to large information files.

The field of automatic indexing and classification has grown to include a variety of methods: (1) statistical methods which use word frequency measures such as those which were first proposed separately by Luhn and Baxendale,[29] (2) positional methods which are based on the position of certain words in titles, or in key sentences, or in a specific relationship to prepositional forms, (3) assignment methods which are based on a dictionary of included or excluded terms as well as key terms such as "summary" or "in conclusion."[30]

---

[24]Mary E. Stevens. AUTOMATIC INDEXING: A STATE-OF-THE-ART REPORT, U.S. Bureau of Standards, Monograph #91, (March 30, 1965), 290 p.; reissued with additions and corrections, February, 1970; the revision contains over 800 new bibliographic items.

[25]S. F. Dennis. "The Design and Testing of a Fully Automatic Indexing-Searching System for Documents Consisting of Expository Text." INFORMATION RETRIEVAL -- A CRITICAL VIEW, edited by G. Schecter, based on the Third Annual Colloqium on Information Retrieval, May 12-13, 1966, Thompson Books, (1967), pp. 67-94.

[26]C. David Batty. "Automatic Generation of Index Languages." JOURNAL OF DOCUMENTATION, 25:2 (June, 1969), pp. 142-151. P. E. Jones points out that the statistical methods suggested for automatic documentation are similar to those used in psycholinguistics and content analysis in his article, "Historical Foundations of Research on Statistical Association Methods for Mechanized Documentation," in STATISTICAL ASSOCIATION METHODS FOR MECHANIZED DOCUMENTATION, edited by M. E. Stevens, V. E. Guiliano and L. B. Heilprin, Symposium proceedings, March 17-19, 1964, U.S. Government Printing Office, U.S. Bureau of Standards, Misc. Publication, #269, (December 15, 1965), pp. 3-8; In a later article written with R. M. Curtice, "A Framework for Comparing Term Association Measures," AMERICAN DOCUMENTATION, 18:3 (July, 1967), pp. 153-161, he notes the overall similarity of such measures and their relation to the 2x2 contingency tables often used for two group comparisons.

[27]H. E. Stiles. "The Association Factor in Information Retrieval," JACM, 8 (April, 1961), pp. 271-279; and "Progress in the Use of the Association Factor in Information Retrieval," in PROCEEDINGS OF A SYMPOSIUM ON MATERIALS INFORMATION RETRIEVAL, Dayton, Ohio: Air Force Materials Laboratory, November 28-29, 1962, Technical Document ASD-TDR-63-445, (May, 1963), pp. 143-153.

[28]Lauren B. Doyle. "Breaking the Cost Barrier in Automatic Classification." System Development Corporation, Report #Sp-2516, (July 1, 1966), 62 p.

[29]H. P. Luhn. "A Statistical Approach to Mechanized Encoding and Searching of Literary Information." IBM JOURNAL OF RESEARCH AND DEVELOPMENT, 1 (1957), pp. 309-317. Phyllis Baxendale. "Machine-made Index for Technical Literature--An Experiment." IBM JOURNAL OF RESEARCH AND DEVELOPMENT, 4:2 (October, 1958), pp. 355-361.

[30]P. Zunde summarizes a wide variety of each type of research in progress as he presents his own Formal Automatic Indexing of Scientific Texts (FAST) system; see AUTOMATIC INDEXING OF MACHINE READABLE ABSTRACTS OF SCIENTIFIC DOCUMENTS, Bethesda, MD., Documentation, Inc., Report #AFOSR65-1425, (September, 1965), 213 p. J. E. Rush, R. Salvador and A. Zamora. "Automatic Abstracting and Indexing. II. Production of Indicative Abstracts by Application of Contextual Inference and Syntatic Coherence Criteria." JASIS, (July-August, 1971), 22:4, pp. 260-274. The authors summarize the previous work of H. P. Edmundson, "New

Because the state of the art has been summarized by Stevens, Batty, Borko, and Zunde, as well as other researchers, this chapter will only outline the developments related to the computer and subject indexing and offer summaries of those experiments of particular relevance to the present study. A long sought goal of automatic indexing has been the creation of software programs which would function as well as, or in a manner similar to, the functioning of the human mind. Such procedures are very different from the usual methods of searching, classifying, or indexing as Bush has observed:

> "The real heart of the matter of selection, however, goes deeper than a lag in the adoption of mechanisms by libraries, or a lack of development of devices for their use. Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing. When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. ... The human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thought, in accordance with some intricate web of trials carried by the cells of the brain... Man cannot hope to duplicate this mental process artificially, but he certainly ought to be able to learn from it."[31]

The seeking after "some intricate web" of associations by programs was not the beginning. The first use of the machine in indexing was the manipulation of index entries previously selected by human indexers.[32] Clapp has observed that the discovery of punched cards began a whole new approach to content analysis:

> "It was suddenly realized, for one thing, that the punched card, so far from being a mere transcript of a book-keeper's ledger on a census form, represented a very respectable intellectual exercise, one involving the logic of classes. It now became possible to think of content analysis in terms of the intersection of little circles, and "co-ordinate indexing," led by the uniterm system and making a virtue of necessity, offered the prospect that grammar might now be dispensed with and the world be analyzed into elemental or atomic concepts and recombined at will."[33]

Mortimer Taube studied existing indexing systems for the Armed Forces Technical Information Agency and then created the "uniterm" system which sought to avoid the problems of heading order altogether by allowing word associations throughout the system. Taube's system, not to be confused with later systems under similar names, was one of the early formalizations of an indexing system based on words extracted from text.[34]

Methods in Automatic Extracting," JACM, 16:2, (April, 1969), pp. 264-285. Edmundson and his associates developed four methods of sentence selection: Cue, Key, Title, and Location methods. The key method utilizes Luhn's frequency of occurrence ideas. The location method utilizes methods similar to those of Baxendale. The cue method makes use of a list of bonus words, stigma words, and null words to weigh sentences. The title method uses a glossary of title and subtitle words as a means of adding positive weight to sentences which contain words in that glossary. Edmundson found that all these methods improve the process of sentence selection, except the key method. Rush, Salvador, and Zamora have experimented with methods whose main contribution lies in the area of selecting sentences for exclusion from the abstract. They utilize a dictionary called the WORD CONTROL LIST containing an alphabetically ordered set of words and phrases. Each textual sentence is checked against this dictionary to find its semantic weight and its syntatic value. The result of this process is sentence retention or deletion.

[31]Vannevar Bush. "As We May Think." ATLANTIC MONTHLY, 176 (July, 1945), pp. 101-108; later part of his ENDLESS HORIZONS, Public Affairs Pamphlet, (1946), 182 p.

[32]Jesse H. Shera. "Mechanical Aids in College and Research Libraries." AMERICAN LIBRARY ASSOCIATION BULLETIN, 32 (1938), pp. 818-819. E. M. Fair. "Inventions and Books--What of the Future?" LIBRARY JOURNAL, LXI (1938), pp. 47-51.

[33]Verner A. Clapp. "Research in Problems of Scientific Information--Retrospect and Prospect." AMERICAN DOCUMENTATION, 14:1 (January, 1963), pp. 1-9.

[34]Mortimer Taube and I. S. Wachtel. "The Logical Structure of Coordinate Indexing." AMERICAN DOCUMENTATION, 4:1 (April, 1953), pp. 67-88. "Functional Approach to Bibliographic Organization; A Critique and a Proposal." BIBLIOGRAPHIC ORGANIZATION, edited by J. H. Shera and M. E. Egan. University of Chicago Press, (1951), pp. 57-71.

Because titles of documents are relatively easy to identify and process, various systems have been developed for using title words as substitutes for created subject headings. Ohlman is generally credited with the development of permuted indexes.[35] Such indexes under the names "KWIC" and "KWOC" have grown to be standard tools in many fields. Questions still remain concerning the adequacy of title words in representing the contents of documents. Feinberg has recently completed a major study in this area,[36] and Vickery offers a summary of recent research on this aspect of the question.[37] Both authors conclude that unaided title extracting methods have gone as far as they can go. Noninformative titles and specific user needs will require human intervention and selection.

Luhn proposed that function words be deleted and the remaining text words be statistically analyzed as a means of creating abstracts finding indexing terms.[38] Fasana notes that the reaction to rigidly structured indexing schemes with controlled vocabularies led to a return to the single word as the basic indexing unit.[39] Nevertheless, the major problems of synonymy, ambiguity, and the lack of a syndetic apparatus soon forced a reassessment of the idea of "single word" indexing. With working experience it became obvious that terms should not only be bound together, but also that various indicators of relationship were needed. "Roles" and "links" were developed:

> "The doctrine of equal values for terms was not maintained, and hierarchical relationships were introduced. It was not possible to dispense with subject authority and cross-reference systems. Even the assumed advantage of free coordination and the almost unlimited possibility for the combination of terms turned out, in many cases, to be disadvantages requiring changes in the structure of headings. Terms had to be linked in order to prevent undesirable coordination. role indicators were introduced to serve as standard subdivisions or modifiers and polyterms or 'concepts' were created to reduce false drops..."[40]

With all of the problems found in conventional alphabetic subject arrangement, including word order in headings, forms for phrase headings, and subdivisions, it "still remains to be proved that any of the newer systems meet the needs of a large collection as well as, or any better than, alphabetic subject headings."[41] Bernier has criticized text derivative indexing maintaining that the important things to be indexed are the concepts which the words in a text only symbolize.[42] Swanson has also noted that problems arise because the same concept may be expressed with different words; he insists that the size

---

[35]H. Ohlman. "Permutation Indexing: Multiple-Entry Listings on Electronic Accounting Machines." System Development Corporation, (November 5, 1957), unpublished. Luhn and Rocketdyne Corporation were also working on this problem.

[36]Hilda Feinberg. A COMPARATIVE STUDY OF TITLE DERIVATIVE INDEXING TECHNIQUES. Columbia University School of Library Service, D. L. S. Thesis, (1972), 390 p.

[37]Brian C. Vickery. "Document Description and Representation." ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY, 5. Encyclopedia Britannica, (1971), pp. 113-140; especially, pp. 118-119.

[38]H. P. Luhn, op. cit. See also the collection of his works edited by Claire K. Shultz. H. P. LUHN: PIONEER OF INFORMATION SCIENCE, SELECTED WORKS. Spartan Books, (1968), A.S.I.S., 320 p.

[39]Paul Fasana. "A Definition of Indexing." TUTORIAL SESSIONS ON INDEXING, edited by B. Flood for the New York Chapter of A.D.I. Drexel Press, Drexel Library School Series, #24, pp. 1-43.

[40]Susan Artandi and Theodore C. Hines. "Roles and Links, or Forward to Cutter." AMERICAN DOCUMENTATION, 14:1 (January, 1963), pp. 74-77. See also B. Montague, "Testing, Comparison, and Evaluation of Recall, Relevance, and Cost of Coordinate Indexing with Links and Roles." AMERICAN DOCUMENTATION, 16:3 (July, 1965), pp. 201-208.

[41]Bella E. Schachtman. "Subject Indexing Mythology." LIBRARY RESOURCES AND TECHNICAL SERVICES, 8:3 (Summer, 1964), pp. 236-247; quote is from page 237.

[42]Charles L. Bernier. "Subject Index Production." LIBRARY TRENDS, 16:3 (January, 1968), pp.389-397. See also C.L. Bernier and Evan J Crane, "Correlative Indexes VIII; Subject Indexing vs. Word Indexing." JOURNAL OF CHEMICAL DOCUMENTATION, 2:2 (April, 1962), pp. 117-122.

of a language sample processed for automatic indexing is a crucial problem.[43]

On the other hand, Spiegal defends the basic assumption of automatic indexing:

> "We assume that the information contained in a message is carried by the words that make it up and by the manner in which they are strung together. Further we assume that a person generating a message... chooses words in a non-random fashion and combines them according to semantic and syntactic rules that are regular and, at least in our culture, predictable."[44]

Moss has also noted the crucial importance of words in indexing:

> "We are, in fact, always indexing words, and it makes no difference whether the words symbolize documents or subjects in them . . . only words can be indexed . . . for a long time to come the only symbols which all of the different specialists and non-specialists have in common and in which there is any sort of agreement for indexing, storage, and retrieval are the words of everyday speech and their printed equivalents."[45]

Having stated that all indexing must, of necessity, deal with words taken from text or with words which verbalize text concepts, serious problems remain concerning the relationship of words to the concepts they represent. Any research effort which utilizes words in text has to deal with the problem of word meanings. Williams summarizes the problem:

> "The essence of the retrieval problem is that some concepts are referred to by more than one term, and some term refer to more than one concept. Thus, the multiple meanings cause both false hits and missing true hits."[46]

Preschel has investigated the relationship between conceptual agreement among indexers and actual term selection agreement among those same indexers.[47] She found that the previous measures of indexer consistency did not distinguish term consistency and concept consistency. Her study indicates that while concept agreement was relatively high, agreement about the words to represent concepts was much lower.

Borko illustrates the word problem with the statement, "Our Host turned on the barbeque spit." He points out that this phrase may be interpreted in numerous ways depending on the context.[48] Wyllys differentiates between "polysemic ambiguity," which refers to the fact that a word or group of words may have more than one meaning, and "string ambiguity," which is created by attaching one word to another as in the case of "scientific information handling."[49] Weiss[50] has explored the

---

[43]Donald R. Swanson. "Research Procedures." MACHINE INDEXING: PROGRESS AND PROBLEMS, op. cit., pp. 281-304; and "Searching Natural Language Text by Computer." SCIENCE, 132 (1960), pp. 1099-1104.

[44]Joseph Spiegal and Edward Bennett. "A Modified Statistical Association Procedure for Automatic Document Content Analysis and Retrieval." STATISTICAL ASSOCIATION METHODS FOR MECHANIZED DOCUMENTATION, op. cit., pp. 47-60; quote is from p. 47.

[45]R. Moss. "Minimum Vocabularies in Information Indexing." JOURNAL OF DOCUMENTATION, 23:3 (September, 1967), pp. 176-199; quote is from p. 183. James L. Dolby and Howard L. Resnikoff are also defenders of the necessity of studying words, see their "On the Structure of Written English Words." LANGUAGE, 40:2 (April-June, 1964), pp. 167-196; especially p. 167.

[46]John H. Williams, Jr. "Functions of a Man-Machine Interactive Information Retrieval System." JASIS, 22:5 (September-October, 1971), pp. 311-317; quote is from p. 316.

[47]Barbara Preschel. "Indexer Consistency in Perception of Concepts and in Choice of Terminology." D. L. S. Study in progress, Columbia University School of Library Service, (1972).

[48]Harold Borko. "Automatic Indexing Process." TUTORIAL SESSIONS ON INDEXING, op. cit., p. 121.

[49]R. E. Wyllys. "Extracting and Abstracting by Computer." AUTOMATED LANGUAGE PROCESSING, op. cit., pp. 127-179; especially p. 169.

[50]S. F. Weiss. "Automatic Resolution of Ambiguities from Natural Language Text." REPORT ON ANALYSIS, DICTIONARY CONSTRUCTION, USER FEEDBACK, CLUSTERING, AND ON-LINE RETRIEVAL, (henceforth ISR-18). Cornell University, Report ISR-18, (October, 1970), pp. IV-2ff.

problems of words with multiple meanings for the SMART system.[51] He notes that Dimsdale and Lamson[52] found, when they examined the word "cell," that in a specific field such ambiguities are not as great a problem as expected.

Fairthorne contrasts the way in which a reader utilizes the words of a document "to find out what someone has said," with the way in which an indexer uses those same words to find out not only what has been said but how what has been said will interest the kinds of readers he serves.[53] Thus what a document is "about" depends on who is reading the words and for what purposes. Maron sums up the language problem:

> "Even when language is used primarily to convey information, we find that it is often vague, ambiguous, imprecise, changeable, idiomatic, and, above all, exceedingly complex...the most amazing aspect of language is the fact that in spite of its vagueness, ambiguity and imprecision, human beings are able to use it with success."[54]

The word problem is further complicated by the limitations of computers and computer programming languages. All computer programs can read only the "form" of a word. Crane and Bernier summarize the machine problem:

> "Machines can not think, but they can tirelessly do many things with great rapidity and accuracy. Machines can not provide nonnumerical information which human beings have not, with forethought, put into them."[55]

As coordinate indexing became entangled in the ambiguities of our language, other means of automatic indexing going beyond single words were sought out. Doyle has discussed the technical problems of automatic classification and noted that disillusionment with single word coordinate indexing caused a renewed interest in using the computer to generate new kinds of indexes.[56]

Following Luhn, Maron and Kuhns published a paper entitled, "On Relevance, Probablistic Indexing and Information Retrieval,"[57] which introduced probability techniques into the field of automatic classification. They used 405 abstracts to produce categories containing sets of manually selected keywords. Edmundson and Wyllys took issue with the idea of absolute or "raw" frequencies used by Luhn.[58] Instead they proposed a measure of word significance based upon the frequencies of words compared with the frequency of words in general. Their method compares frequencies of words from a particular set of documents with the frequencies of those same words in a larger corpus of words.[59]

---

[51]Gerard Salton and M. E. Lesk. "Computer Evaluation of Indexing and Text Processing." JACM, 15:1 (January, 1968), p. 8-36.

[52]B. Dimsdale and B. C. Lamson. "A Natural Language Information Retrieval System." PROCEEDINGS OF THE I.E.E.E., 54:12 (December, 1966), pp. 1636-1640.

[53]Robert A. Fairthorne. "Content Analysis, Specification and Control." ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY, 4, edited by Carlos A. Cuadra. Encyclopedia Britannica, (1969), pp. 73-109. See also his "Functional Analysis of Information Retrieval." U. S. GOVERNMENT RESEARCH AND DEVELOPMENT REPORTS, 69:20 (January 10, 1969) AD 677 289.

[54]M. E. Maron. "A Logician's View of Language-Data Processing." NATURAL LANGUAGE AND THE COMPUTER, edited by Paul L. Garvin. McGraw-Hill, (1963), pp. 128-150; quote is from p. 136.

[55]Evan J Crane and C. L. Bernier. "Indexing and Index Searching." PUNCH CARDS, THEIR APPLICATION TO SCIENCE AND INDUSTRY, edited by R. S. Casey and J. W. Perry. Reinhold, second edition, (1958), pp. 510-527. A recent article by J. Nievergelt and J. C. Farrar, "What Machines Can and Cannot Do," COMPUTING SURVEYS, 4:2, (June, 1972), pp. 81-96, points out the continuing argument about machine capacities including: Can a machine think?, Can a Machine reproduce itself?, and, Are there tasks we can prove no machine will ever be able to perform?

[56]Lauren B. Doyle. "Is Automatic Classification a Reasonable Application of the Statistical Analysis of Text?" JACM, op. cit., pp. 473-489.

[57]M. E. Maron and J. L. Kuhns. "On Relevance, Probablistic Indexing and Information Retrieval." JACM, 7:30 (July, 1960), pp. 216-244.

[58]H. P. Edmundson, and R. E. Wyllys. "Automatic Abstracting and Indexing--Survey and Recommendations." JACM, 4:5 (May, 1961), pp. 226-234.

[59]Harriet R. Meadow. "Statistical Analysis and Classification of Documents." COMMUNICATIONS OF THE ACM, 4:5 (1961). Here Meadow reports her computer evaluation of the Edmundson and Wyllys proposal in IRAD task #0353 for the Federal Systems Office of IBM.

As research moved beyond single words to various combinations of word sets, questions were immediately raised about the relationships among words, the number of words to be included, and the ordering of words which are to appear together either in subject headings or in classification groups. Vickery has noted that the interests in word relationships moves us closer to the allied interests of classification and subject heading research.[60]

Numerous experiments have been conducted to determine word groupings through "association maps," "clumps," "clusters," and "factor analysis." These experiments are discussed here because they bear a close relationship to the problems of multiterm or phrase headings.

In 1958, Tanimoto formulated the problem of classification in terms of attributes of matrix functions.[61] In 1961, Borko applied this procedure to a set of 997 psychological abstracts.[62] Later he expanded this procedure to deal with larger matrices in a more efficient manner.[63] Arnovick, Liles and Wood have reviewed the various techniques for factor analysis.[64] Difondi analyzed ninety-four documents into thirty-nine different factors and then classed documents on the basis of those factors.[65]

Parker-Rhodes, Needham, and Sparck Jones reject factor analysis because of the large data matrices required. Their "clumping" method seeks out words which strongly co-occur with a given text word "x." The process differs from that proposed by Maron in two ways: (1) categories are not mutually exclusive; word may belong to several clumps, and (2) there are no predefined limits on the categories of terms. Sparck Jones has elaborated a series of clump types including "strings" and "stars."[66] In the United States, Dale and Dale have applied the "clump" idea in their work at the University of Texas.[67]

Baker[68] has utilized the later class analysis theory of Lazarsfeld[69] to create an automatic classification method computed on the basis of the probability that a document containing a certain pattern of keywords belongs to a certain class. Winter has also used a modified form of the same idea.[70] Williams and Hillman use a multidimensional structure, as contrasted with a two dimensional array, for their experiments with automatically generated

---

[60]Brian C. Vickery. "Developments in Subject Indexing." JOURNAL OF DOCUMENTATION, 11:1 (March, 1955), p. 1-11.

[61]T. T. Tanimoto. "An Elementary Mathematical Theory of Classification and Prediction." IBM, (1958), 10 p.

[62]Harold Borko. "The Construction of an Empirically Based Mathematically Derived Classification System," Report No. Sp-585. System Development Corporation, (October 26, 1961), 23 p. See also Harold Borko and M. D. Bernick, "Automatic Document Classification," Technical Memo #TM-771, (November, 15, 1962); and "Part II, Additional Experiments," Technical Memo #TM-771/001/00, (October 18, 1963), 33 p.

[63]Harold Borko. "Indexing and Classification." AUTOMATED LANGUAGE PROCESSING, op. cit., pp. 99-125.

[64]G. W. Arnovick, J. A. Liles, and J. S. Wood. "Information Storage and Retrieval: Analysis of the State-of-the-Art." JOURNAL OF THE SPRING JOINT COMPUTER CONFERENCE. AFIPS Press, (1964), pp. 537-561.

[65]N. M. Difondi. "Statistical Information Retrieval System." Griffis Air Force Base, New York, Rome Air Development Center, October, (1969), 56 p.

[66]A. F. Parker-Rhodes and R. M. Needham. "The Theory of Clumps." Cambridge, England: Cambridge Language Research Unit, Report #ML 126, (February, 1960). Karen Sparck Jones and D. J. Jackson. "Current Approaches to Classification and Clump-Finding at the Cambridge Language Research Unit." COMPUTER JOURNAL, 10 (May, 1967), pp. 29-37.

[67]A. G. Dale and N. Dale. "Some Clumping Experiments for Associative Document Retrieval." AMERICAN DOCUMENTATION, 16:11 (January, 1965), pp. 5-9.

[68]F. B. Baker. "Information Retrieval Based on Latent Class Analysis." JACM, 9:4 (October, 1962), pp. 512-521.

[69]P. F. Lazarsfeld. "Latent Structure Analysis." THE AMERICAN SOLDIER, edited by S. A. Stouffer. Princeton University Press, (1950), chapters 10 and 11.

[70]W. K. Winter. "A Modified Method of Latent Class Analysis for File Organization in Information Retrieval Work." JACM, 12:3 (July, 1965), pp. 356-363.

classification.[71] The use of discriminant analysis is illustrated by Williams' study of a small set of reference documents previously classified by human indexers. He derives theoretical frequencies for each word type on the basis of this sample and then calculates both the word dispersion within each category and between categories. The results of these calculators are a set of weighted coefficients for a set of multiple discriminant functions.[72]

Doyle utilized Ward's hierarchical grouping procedure to create "association maps" for document classification.[73] Later he revised his procedures to make the cost of processing large files more economic.[74] Similar procedures have been used by Stiles and Dennis.[75] Stiles studied 100,000 documents indexed by the uniterm system. His strategy was "to generate by machine an expanded list of request terms that will serve as a bigger net to catch documents."[76] This strategy provided listings of term associations based on frequencies of term cooccurrence utilizing the 2x2 contingency tables for CHI-square. Meetham utilized a similar association measure for text analysis.[77] Oswald went beyond word pairs to create similar word groups of up to six words. These high frequency cooccurring word groups were then used for "extracts" and as index terms.[78]

Stone and Rubinoff sought word groupings that tended to "cluster" in a few documents. Their sample of 70,000 words from 217 reviews in COMPUTING REVIEWS was examined for technical or specialty words and then a Poisson probability test was utilized to select the most likely clusters of words.[79] Bonner, Johnson and Lafluente, Mooers, and Vaswani have all reported research utilizing clusters or "lattice" structures.[80] Minker, Wilson, and Zimmerman have tested the cluster concept utilizing Salton's SMART clustering system, the Augustson system, and the Zimm Clustering system de-

---

[71]J. H. Williams, Jr. DISCRIMINENT ANALYSIS FOR CONTENT CLASSIFICATION. Bethesda, MD: IBM Corporation, (December, 1965), 272 p. D. J. Hillman. MATHEMATICAL THEORIES OF RELEVANCE WITH RESPECT TO THE PROBLEM OF INDEXING. Lehigh University Center for Information Science, Report #2, (1965), 56 p.

[72]A similar procedure has been utilized by W. G. Hoyle in his report, "Automatic Classification and Indexing: A Supplement." National Research Council of Canada, Radio and Electrical Engineering Division, ERB-793, (November, 1968), 6 p. and appendices.

[73]Lauren B. Doyle. "Semantic Roadmaps for Literature Searchers." JACM, 8:2 (October, 1961), pp. 553-578. J. H. Ward, Jr., and M. E. Hook. "Application of Hierarchical Grouping Procedures to Problems of Grouping Profiles." EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENT, 23 (1963), pp. 69-82.

[74]Lauren B. Doyle. "Breaking the Cost Barrier in Automatic Classification," op. cit.

[75]H. E. Stiles. "The Association Factor in Information Retrieval." JACM, 8:2 (April, 1961), pp. 271-279. S. F. Dennis. "The Construction of a Thesaurus Automatically from a Sample of Text." STATISTICAL ASSOCIATION METHODS FOR MECHANIZED DOCUMENTATION, op. cit., pp. 61-148.

[76]H. E. Stiles. "Machine Retrieval Using the Association Factor." MACHINE INDEXING: PROGRESS AND PROBLEMS, op. cit., pp. 192-205; quote is from p. 192.

[77]A. R. Meetham. "Probabilistic Pairs and Groups of Words in Text." LANGUAGE AND SPEECH, 7:2 (April-June, 1964), pp. 98-106.

[78]V. A. Oswald, Jr. "Automatic Indexing and Abstracting of the Contents of Documents." Planning Research Corporation, (October 31, 1959), prepared for Rome Air Development Center, RADC-TR-59-208, pp. 5-34, and 59-133.

[79]D. C. Stone and M. Rubinoff. "Statistical Generation of a Technical Vocabulary." AMERICAN DOCUMENTATION [illegible]:4 (October, 1968), pp. 411-412.

[80]R. E. Bonner. "On Some Clustering Techniques." IBM JOURNAL OF RESEARCH AND DEVELOPMENT, 8:1 (January, 1964), pp. 22-32. C. N. Mooers. "A Mathematical Theory of Language Symbols in Retrieval." INTERNATIONAL CONFERENCE ON SCIENTIFIC INFORMATION, Area #6 Reports, (1958), pp. 57-94. D. B. Johnson and J. M. Lafuente. "A Controlled Single Pass Classification Algorithm for Multilevel Clustering." ISR-18, op. cit., pp. XII-1 to XII-37. P. K. T. Vaswani. "A Technique of Cluster Emphasis and its Application to Automatic Indexing." PROCEEDINGS OF THE IFIPS CONGRESS, 60, Edinburgh, August 4-10, 1968, Booklet #6. North Holland, (1968), pp. 61-64.

veloped at the University of Maryland. Their test involved the use of the Medlars system and abstracts from the IRE JOURNAL.[81]

In the development and use of all automatic procedures, there is a large element of human intellectual effort as both Richmond[82] and Vickery[83] point out. As early as 1966, Baxendale stated that "statistical models of association . . . which are restricted to frequency of occurrence data have reached the limits of their capacity. Thus far, these models have not been able to establish a warrant for meaningfully relating language 'usage' to frequency of occurrence."[84] Lesk found that no choice of frequency or correlation cutoff point would yield word pairs he considered reliable. He concludes that word-word association measures may be valuable in showing relationships which are not normally apparent and could serve as an aid in dictionary or thesaurus construction as has also been suggested by Stiles.[85] Further he suggests that second order associations of words not associated with each other but both found in association with another word may be helpful in making retrieval more precise.[86]

It is evident that previous research discussed here has concentrated on: (1) analysis of language for human or automatic indexing purposes, and (2) the creation of indexing or search tools such as thesauri or association maps. Adkinson and Stearns, reviewing the use of the computer in the library, suggested three phases of automation in libraries: mechanization of conventional processes, automation of search procedures based on subject matter, and new and different kinds of services based on computer technology. They concluded that, as of 1967, efforts were largely stopped in phase two "because of the difficulty experienced by computers in dealing with natural language and subjective ambiguities."[87]

D. CONCLUDING REMARKS ON THE STATE OF THE ART

In conclusion, the expectation that the computer would somehow magically remove human effort in indexing or find a new way to meaningfully represent the contents of documents to their potential users has floundered on the complexities and ambiguities of user needs and language. Extremely valuable tools such as permuted indexes have been developed through this research.

Handling large information files requires human intellectual effort at several points. In the initial phase, the information file designer's efforts can provide file organization through classification, or through the creation of access points by indexing and an index display. In subsequent phases, the intellectual efforts shift to the user. When confronted by a sequential file, the user must organize a search strategy to search that file in the most efficient way possible.

The perspective of this paper is to give the user of an information system all of the help possible. Thus, this perspective advocates that the information file designer make the effort in the initial phase through file organization, building of indexing vocabularies, and creating alphabetical displays in indexes with adequate cross references. Computers and computer programs are seen as means of assisting information system designers in these efforts. Computers have proved to be useful tools in such design work. A

---

[81]Jack Minker, Gerald A. Wilson, and Barbara H. Zimmerman. EVALUATION OF QUERY EXPANSION BY THE ADDITION OF CLUSTERED TERMS, FOR A DOCUMENT RETRIEVAL SYSTEM. University of Maryland Computer Science Center, (October, 1971), 97 p.

[82]Phyllis A. Richmond. "Transformation and Organization of Information Content: Aspects of Recent Research in the Art and Science of Classification." PROCEEDINGS OF THE 31ST ANNUAL MEETING AND CONGRESS OF THE INTERNATIONAL FEDERATION FOR DOCUMENTATION, (FID), October 7-16, 1965, Washington, DC. Spartan Books, (1966), Volume 2, pp. 87-106; see especially p. 95.

[83]Brian C. Vickery. "Document Description and Representation," op. cit., p. 122.

[84]Phyllis Baxendale. "Content Analysis, Specification, and Control." ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY, I, edited by Carlos A. Cuadra. Interscience, (1966), pp. 71-106; quote is from p. 96.

[85]M. E. Lesk. "Word-Word Association in Document Retrieval Systems." AMERICAN DOCUMENTATION, 20:1 (January, 1969), pp. 27-38. H. E. Stiles. "The Association Factor in Information Retrieval," op. cit.

[86]W. S. Cooper. "On Higher Level Association Measures." JASIS, 22:5 (September-October, 1971), pp. 354-355.

[87]Burton W. Adkinson and C. M. Stearns. "Libraries and Machines--a Review." AMERICAN DOCUMENTATION, 18:3 (July, 1967), pp. 121-124; quote is from p. 124.

number of examples are found on the following pages.

Research is continuing in the direction of more efficient handling of large files, and more complex pattern matching programs, seeking to represent word context by computer programs. Title and text derivative methods, such as KWIC or KWOC techniques, have become almost standard tools as early alerting devices. Also in the area of computer-assisted indexing, the computer has proved useful for manipulating text, handling a variety of printing formats, and providing multiple access points from single bibliographic entries to which indexing terms have been assigned by human indexers.

Hines, Harris, and Colverd have reported a system of programs for computer-assisted indexing developed at Columbia University School of Library Service. These programs allow "one time only" creation of bibliographic entries in machine readable form with manually selected subject headings. This converted entry is machine-expanded so that author, title, and subject access points are created, a dictionary catalog output is organized alphabetically, and the index is produced in a page and column format which is similar to the H. W. Wilson Company indexes. Such organization of the index entries avoids double look-up during searching.[88]

The computer has also been connected to a variety of display devices and utilized as a means of recording, storing, and retrieving information such as: thesaurus listings, records of index term usage, and suggested index term candidates. Bern describes a system which permits a list of candidate terms to be displayed after they have been derived from textual material.[89] Bennett has designed a "Negotiated Search Facility" which allows an indexer to utilize a display station to search IBM library documents in a variety of ways.[90] Other such systems include BOLD,[91] DIALOG,[92] AUDACIOUS,[93] and the above mentioned SMART system. Thompson has recently reviewed the literature of such interactive, display station oriented information systems.[94] Artandi includes discussion of such various computer derived or displayed information operations in her "Document Description and Representation."[95]

The text processing, text searching, and text formatting possibilities of the computer connected with a remote display device are tremendous. Most of the calculations, permutations, tables, charts and matrices of the author's dissertation[96] were made possible by the powerful remote terminal system of the Columbia University Computer Center.

Several "higher level" programming languages allow for textual manipulation of large, variable length character strings.

---

[88]Theodore C. Hines, Jessica L. Harris, and Martin Colverd. "Experimentation with Computer-Assisted Indexing." JASIS, 21:6, (November-December 1970), pp. 402-405.

[89]G. M. Bern. "Description of FORMAT, a Text Processing Program." COMMUNICATIONS OF THE ACM, 12 (March, 1969), pp. 141-146.

[90]J. L. Bennett. "On-line Access to Information; NSF as an Aid to the Indexer/Cataloger." AMERICAN DOCUMENTATION, 20:3 (July, 1969), pp. 213-220.

[91]H. P. Burnaugh. THE BOLD USERS' MANUAL. System Development Corporation, SDC TM-2306/004/01, (January, 1967).

[92]R. K. Summit. "An Operational On-Line Reference Retrieval System." PROCEEDINGS OF THE ACM NATIONAL MEETING 1967, DC: Thompson Book Co., (1967), pp. 51-56.

[93]R. R. Freeman and P. Atherton. AUDACIOUS--AN EXPERIMENT WITH AN ON-LINE INTERACTIVE REFERENCE RETRIEVAL SYSTEM USING THE UNIVERSAL DECIMAL CLASSIFICATION AS THE INDEX LANGUAGE IN THE FIELD OF NUCLEAR SCIENCE. American Institute of Physics, UDC Project, (April, 1968), AIP/UCE -7.

[94]D. A. Thompson. "Interface Design for an Interactive Information Retrieval System: A Literature Survey and a Research System Description." JASIS, 22:6 (November-December, 1971), pp. 361-373.

[95]Susan Artandi. "Document Description and Representation." ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY, 6, edited by Carlos Cuadra. Encyclopedia Britannica, (1970), pp. 143-167.

[96]"Computer-assisted Analysis of a Large Corpus of Current Educational Report Vocabulary." D. L. S. dissertation, Columbia University, (1972).

This author has found SPITBOL,[97] designed for high speed pattern matching and text processing, and SNOBOL4[98] relatively easy to learn, conceptually understandable in a nonscientific teaching situation, and capable of extensive matrix and table development.

A problem facing all textual processing for automatic indexing and abstracting, as well as linguistic studies, has been the lack of suitable machine-readable text. More and more index and abstract services are now providing magnetic tape services which can be utilized for this purpose. Once the machine-readable text problem is solved, the remaining problem is storage space. Text processing and word correlation demand tremendous amounts of computer memory. This fact accounts in part for the relatively small textual samples utilized in much research and the amount of study done and re-done on the same samples. Word matching, word frequency tables, lists, or matrices require that the computer have sufficient memory to store, address, and remember the results of varied comparisons on a large scale basis. In many machine configurations one buys increased speed with increased storage; a factor which often does not make for more economical processing.

When the computer is used as a counting, extracting, and formatting tool in the study of language, it is a valuable assistant in the development of better controlled vocabulary indexing languages and indexing aids such as thesauri. It is well to remember that the state of the art remain such that the computer will still do exactly as it is told so that the selection of a language sample for study is not the computer's task. The computer will process whatever data it is given in whatever ways it is instructed. The human intellectual tasks remain:

1) to select a sample of language representative of the written language of the persons in the particular subject field where indexing and abstracting are under consideration;

2) to determine the depth of indexing necessary to serve those persons and their research needs in that specific field before developing specific indexing techniques and computer processing programs; and

3) to plan for systematic feedback from the system users to allow for continuing relevance in the face of changing demands.

The demand that information retrieved be specifically relevant to the interests and purposes of the information system user remains at the heart of the "information exchange" which determines the success or failure of any information system. Even with remote display devices, enhanced programming capacities, and on-line data files organized for rapid retrieval, the dynamics of information exchange system will remain a very human equation that is surprisingly similar to the "reference interviews" conducted by librarians in the past.

---

[97]Robert B. K. Dewar. SPITBOL. Illinois Institute of Technology, (February 12, 1971), Version 2.1.1., v.p.

[98]R. E. Griswold, J. F. Poage, and I. P. Polonsky. THE SNOBOL4 PROGRAMMING LANGUAGE, Second edition. Prentice-Hall, (1971), 256 p.

BIBLIOGRAPHY

Adkinson, Burton W. and Stearns, C. M. "Libraries and Machines--a Review." AMERICAN DOCUMENTATION (henceforth AD), 18:3 (July, 1967), pp. 121-124.

Armitage, Janet E. and Lynch, Michael F. "Articulation in the Generation of Subject Indexes by Computer." JOURNAL OF CHEMICAL DOCUMENTATION, 7:3, (August, 1967), pp. 170-178.

Artandi, Susan. "Document Description and Representation." in ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY. 5, Encyclopedia Britannica, 1970, pp. 143-167.

------. and Hines, Theodore C. "Roles and Links--or Forward to Cutter." AD, 14:1 (January, 1963), pp. 74-77.

Baker, Frank B. "Information Retrieval Based on Latent Class Analysis." JOURNAL OF THE ASSOCIATION FOR COMPUTING MACHINERY, (henceforth JACM), 9:4 (October, 1962), pp. 512-521.

Barhydt, Gordon and Schmidt, Charles T. INFORMATION RETRIEVAL THESAURUS FOR EDUCATION. Case Western Reserve University, 1970. 131 p.

Barr, K. P. "Estimates of the Number of Currently Available Scientific and Technical Periodicals." JOURNAL OF DOCUMENTATION, (henceforth JD), 23:2, (June, 1967), pp. 110-116.

Batty, C. David. "The Automatic Generation of Index Languages." JD, 25:2, (June, 1969), pp. 142-151.

Baxendale, Phyllis B. "Content Analysis, Specification and Control." in ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY, Interscience (1966), pp. 71-106.

------. "Machine-Made Index for Technical Literature--An Experiment." IBM JOURNAL OF RESEARCH AND DEVELOPMENT, 4:2 (October, 1958), pp. 355-361.

Bennett, J. L. "On-line Access to Information: NSF as an Aid to the Indexer/Cataloger." AD, 20:3 (July, 1966), pp. 213-220.

Bern, G. M. "Description of FORMAT, a Text Processing Program." COMMUNICATIONS OF THE ACM, 12 (March, 1969), pp. 141-146.

Bernier, Charles L. "Subject Index Production." LIBRARY TRENDS, 16:1 (January, 1968), pp. 388-397.

------. "Indexing and Thesauri." SPECIAL LIBRARIES, 59, (February, 1968), pp. 98-103.

------. and Crane, Evan J "Correlative Indexes VIII; Subject Indexing vs. Word Indexing." JOURNAL OF CHEMICAL DOCUMENTATION, 2 (1962), pp. 117-122.

Bonner, R. E. "On Some Clustering Techniques." IBM JOURNAL OF RESEARCH AND DEVELOPMENT, 8:1 (January, 1964), pp. 23-32.

Borko, Harold. (ed.). AUTOMATED LANGUAGE PROCESSING. New York: Wiley, 1967, 386 p.

------. "Automatic Indexing Process." in TUTORIAL SESSIONS ON INDEXING, edited by B. Flood, Drexel Press, Library School Series #24, (1967), pp. 121-132.

------. "The Construction of an Empirically Based Mathematically Derived Classification System." Report #SP-585, System Development Corporation, (October 26, 1961), 23 p.

------. and Bernick, M. "Automatic Document Classification." JACM, 10:2 (1963), pp 151-162; and "Additional Experiments." JACM, 11:2 (1964), pp. 138-151.

Bourne, Charles P. METHODS OF INFORMATION HANDLING. Wiley, (1967), 386p.

Burnaugh, H. P. "The BOLD Users' Manual," Santa Monica, California: System Development Corporation, (January, 1967), TM 2306/004/01.

Bush, V. "As We May Think." ATLANTIC MONTHLY. 176 (1945), pp. 101-108.

------. ENDLESS HORIZONS. Public Affairs Press, (1946), 182 p.

Cheydleur, B. F. "Information Retrieval--1966." DATAMATION. 7:10 (October, 1966) pp. 21-25.

Clapp, Verner. "Research in Problems of Scientific Information--Retrospect and Prospect." AD, 14 (January, 1963), pp. 1-9.

Classification Research Group. "Need for a Faceted Classification as a Basis for All Methods of Information Retrieval." LIBRARY ASSOCIATION RECORD. 57:7, (July, 1955), pp. 262-268.

Coates, E. J. "Subject Catalogues: Headings and Structure." Library Association, U.K. (1960), 186 p.

Cooper, W. S. "On Higher Level Association Measures." JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE, (henceforth JASIS), 22:5 (September-October, 1971), pp. 354-365.

Crane, Evan J and Bernier C. L. "Indexing and Index Searching." in PUNCH CARDS: THEIR APPLICATION TO SCIENCE AND INDUSTRY, edited by R. S. Casy and J. W. Perry, Reinhold, (1958), Second Edition, pp. 510-527.

Curtice, R. M. "A Framework for Comparing Term Association Measures." AD, 18:3 (July, 1967), pp. 153-161.

Daily, J. E. "The Grammar of Subject Headings." Columbia University School of Library Service. D. L. S. Thesis, (1957), 222 p.

Dale, A. G. and Dale, N. "Some Clumping Experiments for Associative Document Retrieval." AD, 16:11 (January, 1965), pp. 5-9.

Dennis, S. F. "The Construction of a Thesaurus Automatically from a Sample of Text." in STATISTICAL ASSOCIATION METHODS FOR MECHANIZED DOCUMENTATION, proceedings of a symposium, Washington, DC, May 17-19, (1964), National Bureau of Standards Miscellaneous Publication #269, pp. 61-148.

------. "The Design and Testing of a Fully Automatic Indexing Searching System for Documents Consisting of Expository Text." in INFORMATION RETRIEVAL--A CRITICAL VIEW, edited by G. Schecter, Thompson Books, (1967), pp. 67-94.

Dewar, Robert B. K. SPITBOL, Illinois Institute of Technology, (February 12, 1971) Version 2.1.1. v.p.

Difondi, N. M. "Statistical Information Retrieval System." Rome Air Development Center, Griffis Air Force Base, (October, 1969), 56 p.

Dimsdale, B. and Lamson, B. C. "A Natural Language Information Retrieval System." PROCEEDINGS OF THE IEEE, 54:12 (December, 1966), pp. 1636-1640.

Dolby, J. E. and Resnikoff, Howard L. "On the Structure of Written English Words." LANGUAGE, 40:2 (April-June, 1964), pp. 167-196.

Doyle, Lauren B. "Breaking the Cost Barrier in Automatic Classification." Santa Monica, California: System Development Corporation, (July 1, 1966), Report #SP 2516, 62 p.

------. "Is Automatic Classification a Reasonable Application of the Statistical Analysis of Text?" JACM, 12:4 (October, 1965), pp. 473-489.

------. "Library Science in the Computer Age." System Development Corporation, (December 17, 1959), Report #SP 141, 22 p.

------. "Semantic Roadmaps for Literature Searchers." JACM, 8 (1966), pp. 553-578.

Edmundson, H. P. "New Methods of Automatic Extracting." JACM, 16:2 (April, 1969), pp. 264-285.

------. and Wyllyes, R. E. "Automatic Abstracting and Indexing: Survey and Recommendations." JACM, 4:5 (May, 1961), pp. 226-234.

Fair, E. M. "Inventions and Books--What of the Future?" LIBRARY JOURNAL, 61 (1938), pp. 47-51.

Fairthorne, Robert A. "Content Analysis, Specification, and Control." in ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY, 4, (1969), Encyclopedia Britannica, pp. 73-109.

Farradane, J. E. L. "A Scientific Theory of Classification and Indexing." JD, 6:2 (1950), pp. 83-99; see also certain comments by the same author in 8:2 (1952) pp. 73-92.

Fasana, Paul. "A Definition of Indexing." in TUTORAL SESSIONS ON INDEXING, edited by B. Flood, Drexel Press, Library School Series #24, (1967), pp. 1-43.

Feinberg, Hilda. "A Comparative Study of Title Derivative Indexing Techniques." Columbia University School of Library Service, D. L. S. Thesis, 1971, 390 p. (now in press).

Frarey, Carlyle J. "Subject Heading Revision by the Library of Congress." Columbia University School of Library Service, Masters Thesis, 1951.

------. "A History of Subject Cataloging Principles and Practices in the United States, 1850-1954." Columbia University School of Library Service, D. L. S. Thesis in process.

Freeman, R. R. and Atherton, P. "AUDACIOUS --An Experiment with an On-line Interactive Reference Retrieval System Using the Universal Decimal Classification as the Index Language in the Field of Nuclear Science." American Institute of Physics (AIP), AIP/UDC Project, (April, 1968), AIP, UCE-7.

Gottschalk, C. M. and Desmond, W. F. "Worldwide Census of Scientific and Technical Serials." AD, 14:3 (July, 1963), p. 188.

Griswold, R. E.; Poage, J. F.; and Polansky, I. P. THE SNOBOL4 PROGRAMMING LANGUAGE, Prentice Hall, (1971), 256 p.

Harris, Jessica L. SUBJECT ANALYSIS; COMPUTER IMPLICATIONS OF RIGOROUS DEFINITION. Scarecrow Press, (1970), 279 p.

Hillman, D. J. "Mathematical Theories of Relevance with Respect to the Problem of Indexing." "Report #2, A Algorithm for Document Characterization." Lehigh University Center for Information Science, (1965), 56 p.

Hines, T. C. and Harris, J. L. "The Columbia University School of Library Service System for Thesaurus Development and Maintenance." INFORMATION STORAGE AND RETRIEVAL, 7, (1971), pp. 39-50.

------; Harris, J. L.; and Colverd, M. "Experimentation with Computer-assisted Indexing." JASIS, 21:6 (November-December, 1970), pp. 402-405.

Hoyle, W. G. "Automatic Classification and Indexing: A Supplement." Ottawa: National Research Council of Canada, Radio and Electrical Engineering Division, ERB-793, (November, 1968), 6 p. + Appendices.

Johnson, D. B. and Lafuente, J. M. "A Controlled Single Pass Classification Algorithm for Multilevel Clustering." in REPORT ON ANALYSIS, DICTIONARY CONSTRUCTION, USER FEEDBACK, CLUSTERING, AND ON-LINE RETRIEVAL. Cornell University, Report #LSR-18, (October, 1970), pp. XII-1 to XII-18.

Jones, Paul E. and Curtice, R. M. "A Framework for Comparing Term Association Measures. AD, 18:3 (July, 1967), pp. 153-161.

Kaiser, J. SYSTEMATIC INDEXING. London, Pitmans, (1911), 250 p.

Korotkin, Arthur L.; Oliver, Lawrence H.; and Bergis, D. R. "Indexing Aids, Procedures, and Devices." General Electric Company, Bethesda, MD, Information Systems Operations, (April, 1965), Report # RADC-TR-64-582. AD 616 342, 110 p.

Lazarsfeld, P. E. "Latent Structure Analysis." in THE AMERICAN SOLDIER, edited by S. A. Stouffer, Princeton University Press, (1950), Chapters 10, 11.

Lesk, M. E. "Word-Word Associations in Document Retrieval Systems." AD, 20:1 (January, 1969), pp. 27-38.

Lilley, Oliver. "Terminology, Form, Specificity and the Syndetic Structure of Subject Headings for English Literature." Columbia University School of Library Service, D. L. S. Thesis, (1959), (unpublished).

Luhn, H. P. "The Automatic Creation of Literature Abstracts." IBM JOURNAL OF RESEARCH AND DEVELOPMENT, 2 (April, 1958), pp. 159-165.

------. "A Statistical Approach to Mechanized Encoding for Searching Literary Information." IBM JOURNAL OF RESEARCH AND DEVELOPMENT, 1 (1957), pp. 309-317.

Maron, M. E. "A Logician's View of Language Data Processing." in NATURAL LANGUAGE AND THE COMPUTER, edited by P. Garvin, McGraw-Hill, (1963), pp. 128-150.

------. and Kuhns, J. L. "On Relevance, Probabilistic Indexing and Information Retrieval." JACM, 7 (1960), pp. 216-244.

Meadow, Harriet. "Statistical Analysis and the Classification of Documents." IBM, Federal Systems Division, (1962), IRAD Task # 0353. Also in COMMUNICATIONS OF THE ACM, 4:5, (1961).

Meetham, A. R. "Probabilistic Paris and Groups of Words in a Text." LANGUAGE AND SPEECH, 7:2 (April-June, 1964), pp. 98-106.

Metcalfe, John W. INFORMATION INDEXING AND SUBJECT CATALOGING. Scarecrow, (1957), 338 p.

Minker, Jack; Wilson, Gerald A.; and Zimmerman, Barbara H. "Evaluation of Query Expansion by the Addition of Clustered Terms for a Document Retrieval System." University of Maryland Computer Science Center, (October, 1971), 97 p.

Montague, B. "Testing, Comparison, and Evaluation of Recall, Relevance, and Cost of Coordinate Indexing with Links and Roles." AD, 16:3 (July, 1965), pp. 201-208.

Morris, Jack C. "The Duality Concept in Subject Analysis." Oakridge National Laboratories, (1953), 46 p.

Moss, R. "Minimum Vocabularies in Information Indexing." JD, 23:3 (September, 1967), pp. 179-196.

Nievergelt, J. and Farrar, J. C. "What Machines Can and Cannot Do." COMPUTING SURVEYS, 4:2 (June, 1972), pp. 81-96.

O'Conner, John. "Automatic Subject Recognition." JACM, 12:4 (October, 1965), pp. 490-515.

------. "Some Remarks on Mechanized Indexing and Some Small-scale Empirical Results." in MACHINE INDEXING: PROGRESS AND PROBLEMS, papers presented at the 3rd Institute on Information Storage and Retrieval, (February 13-17, 1961), American University, (1962), pp. 266-279.

Ohlman, H. "Permutation Indexing: Multiple-entry Listing on Electronic Accounting Machines." System Develop-

ment Corporation, (November 5, 1957), (unpublished).

Oswald, V.A., Jr. "Automatic Indexing and Abstracting of the Contents of Documents." Planning Research Corporation, (October 31, 1959), RADC TR 59 208, pp. 5-34 and 59-133.

Parker-Rhodes, A. F. and Needham, R. M. "The Theory of Clumps." Report #Ml 126, Cambridge Language Research Unit (U.K.), (1960), see also ML 139.

Pollard, A. F. C. and Bradford, S. C. "The Inadequacy of the Alphabetical Subject Index." in PROCEEDINGS OF THE SEVENTH CONFERENCE OF ASLIB, (1930), pp. 39-45.

Preschel, Barbara. "Indexer Consistency in the Perception of Concepts and the Choice of Terminology." Special Report for the Office of Education. (In Press).

Prevost, Marie-Louise. "Approach to Theory and Method in General Subject Headings." LIBRARY QUARTERLY, 16:2 (April, 1946), pp. 140-151.

Prywes, Noah S. and Litofsky, B. "All-automatice Processing for a Large Library." Spring Joint Computer Conference, 36, May 5-7, 1970. AFIPS Press, (1970), pp. 323-333.

Ranganathom, S. R. PROLEGOMENA TO LIBRARY CLASSIFICATION. Library Association, U.K., Second Edition, (1957), 487 p.

Richmond, Phyllis A. "Cats: An Example of Concealed Classification in Subject Headings." LIBRARY RESOURCES AND TECHNICAL SERVICES. 3 (Spring, 1959), pp. 102-112.

------. "Transformation and Organization of Information Content: Aspects of Recent Research in the Art and Science of Classification." in PROCEEDINGS OF THE 31ST ANNUAL MEETING OF FID, WASHINGTON, DC, 1965. Spartan Books, (1966), pp. 87-106.

Rush, J. E., Salvador, R. and Zamora, A. "Automatic Abstracting and Indexing. II Production of Indicative Abstracts by Application of Contextual Influence and Syntatic Coherence Critera." JASIS, 22:4 (July-August, 1971), pp. 260-274.

Salton, G. AUTOMATIC INFORMATION ORGANIZATION AND RETRIEVAL. McGraw-Hill, (1968).

------. THE SMART RETRIEVAL SYSTEM: EXPERIMENTS IN AUTOMATIC DOCUMENT PROCESSING. Prentice Hall, (1971), 556 p.; see also the series of ISR reports to the National Science Foundation, various dates.

------. and Lesk, M. E. "Computer Evaluation of Indexing and Text Processing." JACM, 15:1 (January, 1968), pp. 8-36.

Schachtman, B. E. "Subject Indexing Mythology." LIBRARY RESOURCES AND TECHNICAL SERVICES, 8 (Summer, 1964), pp. 236-247.

Schultz, Claire K. (ed.). H. P. LUHN: PIONEER OF INFORMATION SCIENCE: SELECTED WORKS. ASIS, Spartan Books, (1968), 320 p.

Shera, Jesse H. "Mechanical Aids in College and Research Libraries." ALA BULLETIN, 32 (1938), pp. 818-819.

------. "The Sociological Relationship of Information Science." JASIS, 22:2, (March-April, 1971), pp. 76-80.

Sparck Jones, K. and Jackson, P. "Current Approaches to Classification and Cluster Finding at the Cambridge Language Research Unit. COMPUTER JOURNAL, 10 (May, 1967), pp. 29-37.

Spiegel, Joseph and Bennett, Edward. "A Modified Statistical Association Procedure for Automatic Document Content Analysis and Retrieval." in STATISTICAL ASSOCIATION METHODS FOR MECHANIZED DOCUMENTATION, N.B.S. Miscellaneous Publication #269, 1965, pp. 47-60.

Stevens, M. E. AUTOMATIC INDEXING: A STATE-OF-THE-ART REPORT. National Bureau of Standards Monograph #91, (March 30, 1965), reissued with additions and corrections, February, 1970, 290 p.

------., Giuliano, V. E., and Heilprin, L. B. (eds). STATISTICAL ASSOCIATION METHODS FOR MECHANIZED DOCUMENTATION. National Bureau of Standards Miscellaneous publication #269. Proceedings of a Symposium, Washington, DC, March 17-19, 1964, 261 p.

Stiles, H. E. "The Association Factor in Information Retrieval." JACM, 8:2, (April, 1961), pp. 271-279.

------. "Machine Retrieval Using the Association Factor." MACHINE INDEXING: PROGRESS AND PROBLEMS, The 3rd Institute on Information Storage and Retrieval, Washington, DC, February 13-17, 1961, American University, (1962), pp. 192-205.

Stone, D. C. and Rubinoff, M. "Statistical Generation of a Technical Vocabulary." AD, 19:4 (October, 1968), pp. 411-12.

Summit, R. K. "An Operational On-line Reference Retrieval System." in PROCEEDINGS OF THE ACM NATIONAL MEETING, 1967. Thompson Books (1967), pp. 51-56.

Swanson, Donald R. "Research Procedures." in MACHINE INDEXING: PROGRESS AND PROBLEMS, pp 281-304.

Tanimoto, T. T. "An Elementary Mathematical Theory of Classification and Prediction." IBM Corporation, (1958), 10 p.

Taube, Mortimer and Wachtel, I. S. "The Logical Structure of Coordinate Indexing." AD, 4:1 (April, 1953), pp. 67-88.

Tauber, M. F. and Lilley, O. L. "Feasibility Study Regarding the Establishment of an Education Media Research Information Services. Columbia University School of Library Service, (1960), 235 p.

Thompson, D. A. "Interface Design for an Interactive Information Retrieval System: A Literature Survey and a Research System Description." JASIS, 22:6 (November-December, 1971), pp. 361-373.

Vaswani, P. K. T. "A Technique of Cluster Emphasis and Its Application to Automatic Indexing." PROCEEDINGS OF THE IFIPS CONGRESS, 60, EDINBURGH, SCOTLAND, AUGUST 4-10, 1968. Booklet #6, North Holland, (1968), pp. 61-64.

Vickery, Brian C. "Developments in Subject Indexing." JD, 11:1 (March, 1955), p. 1-11.

------. "Document Description and Representation." in ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY, 5, Encyclopedia Britannica, (1971), pp. 113-140.

Ward, J. H. Jr., and Hook, M. E. "Application of Hierarchy Group Procedures to Problems of Grouping Profiles." EDUCATIONAL AND PSYCHOLOGICAL MEASUREMENTS, 23 (1963), pp. 69-82.

Weiss, S. F. "Automatic Resolution of Ambiguities for Natural Language Text." ISR -18, Gerard Salton, Project Director, Cornell University, (1970), Section IV.

Williams, J. H. Jr. "Discriminent Analysis for Content Classification." IBM Corporation, Federal Systems Division, (December, 1965), 272 p. ERIC #ED027917, AD 630 127.

------. "Functions of a Man-Machine Interactive Information Retrieval System." JASIS, 32:5 (September-October, 1971), pp. 311-317.

Winter, W. K. "A Modified Method of Content Classification Analysis for File Organization in Information Retrieval Work." JACM, 12:3 (July, 1965), pp. 356-363.

Wyllys, R. E. "Extracting and Abstracting by Computer." in AUTOMATED LANGUAGE PROCESSING, edited by H. Borko, Wiley, (1967), pp. 127-129.

Zunde, Pranas. "Automatic Indexing for Machine Readable Abstracts of Scientific Documents." Documentation Incorporated, (September, 1965), 213 p. (FAST system report).